**Теоретические основы решения задачи №13**

Алфавит - конечное множество символов. Текст — произвольная последовательность символов данного алфавита. Двоичные тексты. Единицы измерения длины двоичных текстов (бит, байт, производные единицы). Шестнадцатеричное представление двоичных текстов.

Для успешного выполнения данного задания в рамках непрерывного курса информатики у учащихся должны быть сформированы следующие теоретические знания:

1. Мощность алфавита N – это количество символов в этом алфавите.
2. С помощью *i* бит можно закодировать  различных вариантов.
3. Знание таблицы степеней двойки, которая показывает сколько вариантов *N* можно закодировать с помощью *i*  бит:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *N* | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |

1. Знание единиц измерения количества информации:

Минимальной единицей измерения количества информации является бит, а следующей по величине единицей является байт, причем 1 байт = 23 бит = 8 бит.

Компьютер оперирует числами не в десятичной, а в двоичной системе счисления, поэтому в кратных единицах измерения количества информации используется коэффициент 2n. Так, кратные байту единицы измерения количества информации вводятся следующим образом:

1 Кбайт = 210 байт = 1024 байт;  
1 Мбайт = 210 Кбайт = 1024 Кбайт;  
1 Гбайт = 210 Мбайт = 1024 Мбайт;

1 Тбайт = 210 Гбайт = 1024 Гбайт.

1. Алфавитный подход к измерению количества информации.

При алфавитном подходе к определению количества информации отвлекаются от содержания информации и рассматривают информационное сообщение как последовательность знаков определенной знаковой системы.

Чтобы найти информационный объем сообщения (текста) *I*, нужно умножить количество символов L на число бит на символ i: **** .

Рассмотрим решение трех типовых задач.

**Пример 1**

Для регистрации на сайте некоторой страны пользователю требуется придумать пароль. Длина пароля - ровно 10 символов. В качестве символов используются десятичные цифры и 10 различных букв местного алфавита, причем все буквы используются в двух начертаниях: как строчные, так и прописные (регистр буквы имеет значение!).

Под хранением каждого такого пароля на компьютере отводится минимально возможное и одинаковое целое количество байтов, при этом используется посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов.

Определите объем памяти (в байтах), который занимает хранение 25 паролей. В ответе укажите только число.

**Решение:**

Данная задача решается с использованием алфавитного подхода к измерению количества информации.

1. Определим мощность используемого алфавита.

N=10(десятичные цифры)+10\*2(буквы местного алфавита строчные и прописные)=30 символов.

1. Определим количество битов, используемое для хранения одного символа.

По формуле  находим:

N=30 , а 30 не является степенью числа 2.

Т.к. речь идет о целом количестве битов, минимально достаточном для представления одного знака данного алфавита, выбираем ближайшее большее N число, являющееся степенью числа 2.

Получаем N=25, следовательно i=5 бит.

1. Определим информационный вес одного пароля.

Согласно формуле , находим, что бит.

переведем биты в байты: 50/8=7 байтов (округляем результат до целого в большую сторону).

1. Определим количество информации для хранения 25 паролей.

 байт.

**Ответ:** 175 байт.

**Важно:**

1. Десятичные цифры это цифры от 0 до 9 (то есть их 10).
2. Если в задаче используются строчные и заглавные буквы, тогда их количество необходимо умножить на 2.

**Пример 2.**

При регистрации в компьютерной системе каждому пользователю выдается пароль, состоящий из 15 символов и содержащий только символы из 8 символьного набора, A,B,C,D,E,F,G,H. В базе данных для хранения сведений о любых пользователях отводится одинаковое минимально возможное целое число байт.

При этом используют посимвольное кодирование паролей, все символы кодируют одинаковым минимально возможным количеством бит. Кроме пароля для каждого пользователя в системе хранятся дополнительные сведения, для чего выделено целое число байт, одно и то же для всех пользователей. Для хранения сведений о 20 пользователях потребовалось 320 байт.

Сколько байт выделено для хранения дополнительных сведений об одном пользователе?

**Решение:**

1. Определим количество информации, используемое для хранения данных об одном пользователе.

 байт.

1. Для определения количества информации, используемого для хранения пароля одного пользователя воспользуемся формулой:

.

1. Определим количество битов, используемых для хранения одного символа алфавита.

   бит.

1. Теперь определим количество байтов используемое для хранения пароля одного пользователя.

.

1. Тогда количество информации для хранения дополнительных сведений об одном пользователе находим как разность общего количества информации об одном пользователе и количества информации, используемого для хранения пароля одного пользователя.

 байт.

**Ответ:** 10 байтов.

**Важно:**

1. Внимательно прочитать задание: требуется ли округление бит до байтов или нет.

**Пример 3[[1]](#footnote-1)**

В зоопарке 32 обезьяны живут в двух вольерах, А и Б. Одна из обезьян заболела. Сообщение «Заболевшая обезьяна живет в вольере А» содержит 4 бита информации. Сколько обезьян живут в вольере Б?

**Решение:** Для данной задачи необходимо применить формулу: N = 2i. Где i – бит информации,

N – количество вариантов, поэтому информация в 4 бита соответствует выбору одного из 16 вариантов, …

1. … следовательно в вольере А живет 1/16 часть всех обезьян (это **самый важный момент**!)
2. всего обезьян – 32, поэтому в вольере А живет

32/16 = 2 обезьяны

1. поэтому в вольере Б живут все оставшиеся
2. – 2 = 30 обезьян

**Ответ**: 30.

**Важно:**

1. Знать формулу нахождения количества вариантов.

**Решение (вариант 2, использование формулы Шеннона)**:

1. заболевшая обезьяна может жить в вольере А (событие 1) или в вольере Б (событие 2)
2. количество информации в сообщении о произошедшем событии с номером  равно , где  – вероятность этого события; таким образом, получаем вероятность того, что заболевшая обезьяна живет в вольере А:

.

1. у нас не было никакой предварительной информации о том, где живет заболевшая обезьяна, поэтому можно считать, что вероятность определяется количеством обезьян в вольере – если вероятность равна 1/16, то в вольере живет 1/16 часть всех обезьян:

32/16 = 2 обезьяны

1. поэтому в вольере Б живут все оставшиеся
2. – 2 = 30 обезьян

**Ответ**: 30.

**Важно:**

1. Знать формулу Шеннона.

**Пример 4[[2]](#footnote-2) (обратная)**

В корзине лежат 32 клубка шерсти, из них 4 красных. Сколько бит информации несет сообщение о том, что достали клубок красной шерсти?

**Решение (вариант 1)**:

1. красные клубки шерсти составляют 1/8 от всех, …
2. поэтому сообщение о том, что первый вынутый клубок шерсти – красный, соответствует выбору одного из 8 вариантов
3. выбор 1 из 8 вариантов – это информация в 3 бита (по таблице степеней двойки)

**Ответ**: 3.

**Решение (вариант 2, использование формулы Шеннона)** **(обратная).**

1. красные клубки шерсти составляют 1/8 от всех, поэтому вероятность  того, что первый вынутый клубок шерсти – красный, равна 1/8
2. по формуле Шеннона находим количество информации в битах:

бита.

**Ответ**: 3.

**Пример 5[[3]](#footnote-3) (Новое задание)**

В школьной базе данных хранятся записи, содержащие информацию об учениках:

    <Фамилия> – 16 символов: русские буквы (первая прописная, остальные строчные),

    <Имя> – 12 символов: русские буквы (первая прописная, остальные строчные),

    <Отчество> – 16 символов: русские буквы (первая прописная, остальные строчные),

    <Год рождения> – числа от 1992 до 2003.

Каждое поле записывается с использованием минимально возможного количества бит. Определите минимальное количество байт, необходимое для кодирования одной записи, если буквы е и ё считаются совпадающими.

**Решение:**

1. очевидно, что нужно определить минимально возможные размеры в битах для каждого из четырех полей и сложить их;
2. важно! известно, что первые буквы имени, отчества и фамилии – всегда заглавные, поэтому можно хранить их в виде строчных и делать заглавными только при выводе на экран (но нас это уже не волнует)
3. таким образом, для символьных полей достаточно использовать алфавит из 32 символов (русские строчные буквы, «е» и «ё» совпадают, пробелы не нужны)
4. для кодирования каждого символа 32-символьного алфавита нужно 5 бит (32 = 25555), поэтому для хранения имени, отчества и фамилии нужно (16 + 12 + 16)•5=220 бит
5. для года рождения есть 12 вариантов, поэтому для него нужно отвести 4 бита (24 = 16 ≥ 12)
6. таким образом, всего требуется 224 бита или 28 байт

**Ответ** : 28 байт.

1. <http://kpolyakov.spb.ru> [↑](#footnote-ref-1)
2. <http://kpolyakov.spb.ru> [↑](#footnote-ref-2)
3. <http://kpolyakov.spb.ru> [↑](#footnote-ref-3)